On the Randomness of Independent Experiments

Thomas Holenstein[*] and Renato Renner[§]

February 1, 2008

Abstract

Given a probability distribution P_X, what is the minimum amount of bits needed to store a value x sampled according to P_X, such that x can later be recovered (except with some small probability ε)? Or, what is the maximum amount of uniform randomness that can be extracted from x? Answering these and similar information-theoretic questions typically boils down to computing so-called *smooth entropies*. In this paper, we derive explicit and almost tight bounds on the smooth entropies of n-fold product distributions P_X^n.

1 Introduction

1.1 Smooth min- and max-entropy

Smooth min- and max-entropy has been introduced in [RW04, RW05] as a generalization of Shannon entropy. Similarly to Shannon entropy, smooth min- and max-entropy can be used to analyze information-processing tasks such as data compression. However, in contrast to Shannon entropy, which usually only makes sense in an asymptotic setting (where an underlying random experiment is repeated many times), smooth entropies can also be used in the non-asymptotic case.

We start by quickly reviewing the relevant definitions. For the following, let X and Y be random variables with range $\mathcal{X}$ and $\mathcal{Y}$, respectively, and joint distribution P_{XY}. Moreover, for $\varepsilon \geq 0$, let $\mathcal{B}^\varepsilon(P_{XY})$ be the ε-ball of nonnegative functions around P_{XY}, i.e., the set of functions $Q_{XY} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ such that $\|P_{XY} - Q_{XY}\|_1 \leq \varepsilon$, where $\|\cdot\|_1$ denotes the L_1-norm.

Definition 1. *The ε-smooth max-entropy of X given Y is*[1]

$$H_{\max}^\varepsilon(X|Y) := \min_{Q_{XY} \in \mathcal{B}^\varepsilon(P_{XY})} \max_{y \in \mathcal{Y}} \log\big|\operatorname{supp} Q_{XY}(\cdot, y)\big|$$

---

[*]Department of Computer Science; ETH Zürich; Switzerland; `thomahol@inf.ethz.ch`

[§]Department of Applied Mathematics and Theoretical Physics; University of Cambridge; United Kingdom; `r.renner@damtp.cam.ac.uk`

[1]We use $\log(\cdot)$ to denote the binary logarithm.

where $\operatorname{supp} Q_{XY}(\cdot, y)$ denotes the support of the function $Q_{XY}(\cdot, y) : x \mapsto Q_{XY}(x, y)$.

The ε-smooth min-entropy of X given Y is

$$H_{\min}^{\varepsilon}(X|Y) := \max_{Q_{XY} \in \mathcal{B}^{\varepsilon}(P_{XY})} \min_{y \in \operatorname{supp} P_Y} \min_{x \in \mathcal{X}} \log \frac{P_Y(y)}{Q_{XY}(x, y)}$$

where P_Y denotes the marginal distribution of P_{XY}.

The following two statements proven in [RW05] imply that the smooth min- and max-entropies have a (non-asymptotic) operational interpretation. For example, the smooth max-entropy characterizes *data compression*. More precisely, for (x, y) chosen according to P_{XY}, it quantifies the minimum space needed to store x such that, with the help of y, the value x can later be retrieved (except with probability at most ε).

Proposition 1. *Let $\ell_{\mathrm{enc}}^{\varepsilon}(X|Y)$ be the minimum number ℓ such that*

$$\Pr_{(x,y)\leftarrow P_{XY}} \left[d(e(x), y) \neq x \right] \leq \varepsilon$$

for some encoding function $e : \mathcal{X} \rightarrow \{0,1\}^{\ell}$ and some decoding function $d : \{0,1\}^{\ell} \times \mathcal{Y} \rightarrow \mathcal{X}$. Then, for any $0 \leq \varepsilon' < \varepsilon$,

$$H_{\max}^{\varepsilon}(X|Y) \leq \ell_{\mathrm{enc}}^{\varepsilon}(X|Y) \leq H_{\max}^{\varepsilon'}(X|Y) + \log(1/(\varepsilon - \varepsilon')) + 1 .$$

Similarly, the smooth min-entropy characterizes *randomness extraction*. That is, for (x, y) chosen according to P_{XY}, it corresponds to the maximum number of bits that can be computed from x such that these bits are uniformly distributed and independent of y (except with probability ε).

Proposition 2. *Let $\ell_{\mathrm{ext}}^{\varepsilon}(X|Y)$ be the maximum number ℓ such that*

$$\frac{1}{2} \left\| P_{h(X)Y} - P_U \times P_Y \right\|_1 \leq \varepsilon$$

for some extraction function $h : \mathcal{X} \rightarrow \{0,1\}^{\ell}$, where P_U is the uniform distribution on $\{0,1\}^{\ell}$. Then, for any $0 \leq \varepsilon' < \varepsilon$,

$$H_{\min}^{\varepsilon'}(X|Y) - 2 \log(1/(\varepsilon - \varepsilon')) \leq \ell_{\mathrm{ext}}^{\varepsilon}(X|Y) \leq H_{\min}^{\varepsilon}(X|Y) .$$

While, by the above propositions, smooth entropies are directly related to data compression and randomness extraction, they are also useful for the characterization of a variety of other tasks in communication theory and cryptography [RW05, RWW06]. Moreover, they can be generalized to quantum states [Ren05].

1.2 Contributions of this paper

In this paper, we are concerned with the explicit computation of smooth entropies for the case of a finite number of independently repeated experiments. More precisely, we derive the following bounds on the smooth min- and max-entropies $H_{\min}^{\varepsilon}(X^n|Y^n)$ and $H_{\max}^{\varepsilon}(X^n|Y^n)$ of an n-fold product distribution $P_{X^nY^n} := P_{X_1Y_1} \times \cdots \times P_{X_nY_n}$.

Theorem 1. *Let $P_{X^nY^n} := P_{X_1Y_1} \ldots P_{X_nY_n}$ be a probability distribution over $\mathcal{X}^n \times \mathcal{Y}^n$. For any $\delta \geq 0$*

$$H_{\max}^{\varepsilon}(X^n|Y^n) \leq H(X^n|Y^n) + n\delta \ ,$$
$$H_{\min}^{\varepsilon}(X^n|Y^n) \geq H(X^n|Y^n) - n\delta \ ,$$

where $\varepsilon = 2^{-\frac{n\delta^2}{2\log^2(|\mathcal{X}|+3)}}$.

Let $(\mathbf{x}, \mathbf{y})$ be chosen according to the n-fold product distribution $P_{X^nY^n} = P_{X_1Y_1} \ldots P_{X_nY_n}$. The well known asymptotic equipartition property states that

$$\lim_{n\to\infty} \left(\Pr_{\mathbf{x},\mathbf{y}}[P_{X^n|Y^n}(\mathbf{x},\mathbf{y}) \in 2^{-H(X^n|Y^n)\pm n\delta}] \right) = 1$$

for every $\delta > 0$. The main step in order to prove Theorem 1 is to give the following quantitative bound on this convergence.

Theorem 2. *Let $P_{X^nY^n} := P_{X_1Y_1} \ldots P_{X_nY_n}$ be a probability distribution over $\mathcal{X}^n \times \mathcal{Y}^n$. Then, for any $\delta \in [0, \log(|\mathcal{X}|)]$ and $(\mathbf{x}, \mathbf{y})$ chosen according to $P_{X^nY^n}$,*

$$\Pr_{\mathbf{x},\mathbf{y}}\left[-\log(P_{X^n|Y^n}(\mathbf{x},\mathbf{y})) \geq H(X^n|Y^n) + n\delta \right] \leq \varepsilon \ ,$$

and, similarly,

$$\Pr_{\mathbf{x},\mathbf{y}}\left[-\log(P_{X^n|Y^n}(\mathbf{x},\mathbf{y})) \leq H(X^n|Y^n) - n\delta \right] \leq \varepsilon \ ,$$

where $\varepsilon = 2^{-\frac{n\delta^2}{2\log^2(|\mathcal{X}|+3)}}$.

We prove the theorems in Section 2. In Section 3 we show that both these theorems are almost tight (cf. Theorems 3 and 4).

1.3 Related work and proof technique

A bound as in Theorem 2 can be obtained in simpler ways than the one we use. However, we only know of simpler arguments which yield quantitatively weaker bounds.

We sketch two such arguments. The first argument (which only seems to work in case the distributions $P_{X_iY_i}$ are identical) goes as follows: let $Q_{\mathbf{x}|\mathbf{y}}$ be the frequency distribution (i.e., the type) of the pair $(\mathbf{x}, \mathbf{y}) = ((x_1, y_1), \ldots, (x_n, y_n))$. It is well known that $D(Q_{\mathbf{x}|\mathbf{y}} \| P_{X|Y})$ is small with high probability, and an

explicit bound can be given [CT91, Theorem 12.2.1]. The Csiszár-Kullback-Pinsker inequality [CT91, Lemma 12.6.2] then shows that $\|Q_{\mathbf{x}|\mathbf{y}} - P_{X|Y}\|_1$ is small with high probability, which in turn bounds $|H(Q_{\mathbf{x}|\mathbf{y}}) - H(P_{X|Y})|$ by the Fano inequality (more concretely [CT91, Theorem 16.3.2]), and a bound on $|H(Q_{\mathbf{x}|\mathbf{y}}) - H(P_{X|Y})|$ is exactly what we want. This argument shows that the probabilities in Theorem 2 are at most

$$2^{-n \cdot \Theta\left(\frac{\delta^2}{\log^2(|\mathcal{X}|/\delta)} - |\mathcal{X}|\frac{\log(n+1)}{n}\right)} .$$

The major problem with this bound is that it is only useful if $n \gg |\mathcal{X}|$, but additionally the term $\log^2(|\mathcal{X}|/\delta)$ is not tight: it can be strengthened to $\log^2(|\mathcal{X}|)$ as our proof shows (which is interesting if $\delta \ll \frac{1}{|\mathcal{X}|}$).

A different argument is used in [ILL89]. There, the Hoeffding bound is applied on the sum of the independent random variables $\log(1/P_{X|Y}(X_i, Y_i))$. Unfortunately, the Hoeffding bound can only be applied if the random variables have a bounded range, and thus one ignores occurrences where this random variable is very large (which happens with some small probability). Hence, this technique only gives a bound on the second probability in Theorem 2, and this bound is

$$2^{-n \cdot \Theta\left(\frac{\delta^2}{\log^2(|\mathcal{X}|/\delta)}\right)} .$$

As above, this is not tight in case $\delta \ll \frac{1}{|\mathcal{X}|}$.

Instead of using Hoeffding's bound we directly use Chernoff's argument [Che52] which states that an upper bound on $\inf_t e^{-at} M(t)$ for every a suffices for our purpose, where $M(t)$ is the moment generating function of the random variable $\log(1/P_{X|Y}(X, Y))$. In order to make the presentation simpler we do not use Chernoff's theorems explicitly, but instead give the complete proof.

2 Smooth min- and max-entropy of products

2.1 Typical sequences and their probabilities

Lemma 1. *Let P_{XY} be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. Then, for any $t \in \mathbb{R}$ with $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$,*

$$\log\left(\mathrm{E}_{x,y}\left[P_{X|Y}(x,y)^{-t}\right]\right) \leq tH(X|Y) + \tfrac{1}{2}t^2 \log^2(|\mathcal{X}| + 3) ,$$

where the expectation is taken over pairs (x, y) chosen according to P_{XY}.

Proof. For any $t \in \mathbb{R}$, let r_t be the function on the open interval $(0, \infty)$ defined by

$$r_t(z) := z^t - t \ln(z) - 1 . \tag{1}$$

We will use several properties of this function proven in Appendix A.

For any $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, let $p_{x,y} := P_{X|Y}(x,y)$. If $p_{x,y} > 0$ then

$$p_{x,y}^{-t} = r_t\left(\tfrac{1}{p_{x,y}}\right) + t\ln\left(\tfrac{1}{p_{x,y}}\right) + 1 \leq r_t\left(\tfrac{1}{p_{x,y}} + 3\right) + t\ln\left(\tfrac{1}{p_{x,y}}\right) + 1 \ ,$$

where the inequality holds because r_t is monotonically increasing on the interval $[1, \infty)$ (Lemma 3) and $\frac{1}{p_{x,y}} = \frac{P_Y(y)}{P_{XY}(x,y)} \geq 1$. Because $\frac{1}{p_{x,y}} + 3 \in [4, \infty)$ and because r_t is concave on this interval (Lemma 5 which can be applied because $t \in [-\frac{1}{2}, \frac{1}{2}]$), Jensen's inequality leads to

$$\mathrm{E}_{x,y}\left[p_{x,y}^{-t}\right] \leq \mathrm{E}_{x,y}\left[r_t\left(\tfrac{1}{p_{x,y}} + 3\right)\right] + t\,\mathrm{E}_{x,y}\left[\ln\left(\tfrac{1}{p_{x,y}}\right)\right] + 1$$

$$\leq r_t\left(\mathrm{E}_{x,y}\left[\tfrac{1}{p_{x,y}} + 3\right]\right) + t\ln(2)\,\mathrm{E}_{x,y}\left[\log\left(\tfrac{1}{p_{x,y}}\right)\right] + 1 \ .$$

Because $\mathrm{E}_{x,y}[\frac{1}{p_{x,y}}] = \sum_{x,y} P_{XY}(x,y)\frac{P_Y(y)}{P_{XY}(x,y)} = |\mathcal{X}|$ and $\mathrm{E}_{x,y}[\log(\frac{1}{p_{x,y}})] = H(X|Y)$, we obtain

$$\mathrm{E}_{x,y}\left[p_{x,y}^{-t}\right] \leq r_t(|\mathcal{X}| + 3) + t\ln(2)H(X|Y) + 1 \ .$$

Furthermore, because $\log(a) \leq \frac{1}{\ln(2)}(a - 1)$,

$$\log\left(\mathrm{E}_{x,y}\left[p_{x,y}^{-t}\right]\right) \leq \tfrac{1}{\ln(2)}r_t(|\mathcal{X}| + 3) + tH(X|Y) \ .$$

Finally, together with Lemma 6, since $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$, we conclude

$$\log\left(\mathrm{E}_{x,y}\left[p_{x,y}^{-t}\right]\right) \leq \left(\tfrac{1}{\ln(2)} - 1\right)t^2 \log^2(|\mathcal{X}| + 3) + tH(X|Y) \ .$$

The assertion follows because $\frac{1}{\ln(2)} - 1 \leq \frac{1}{2}$. $\qquad\qquad\square$

Lemma 2. *Let P_{XY} be a probability distribution and let γ be the function on $\mathcal{X} \times \mathcal{Y}$ defined by*

$$\gamma(x,y) := -\log(P_{X|Y}(x,y)) - H(X|Y) \ .$$

Then, for any $t \in \mathbb{R}$ with $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$,

$$\mathrm{E}_{x,y}\left[2^{t\gamma(x,y)}\right] \leq 2^{\frac{1}{2}t^2 \log^2(|\mathcal{X}|+3)} \ .$$

Proof. The assertion follows directly from Lemma 1, that is,

$$\mathrm{E}_{x,y}\left[2^{t\gamma(x,y)}\right] = 2^{-tH(X|Y)}\,\mathrm{E}_{x,y}\left[P_{X|Y}(x,y)^{-t}\right]$$

$$\leq 2^{-tH(X|Y)} \cdot 2^{tH(X|Y) + \frac{1}{2}t^2 \log^2(|\mathcal{X}|+3)} \ . \qquad\qquad\square$$

Proof (of Theorem 2). Let $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, and let γ_i be the function defined in Lemma 2 for the probability distribution $P_{X_i Y_i}$. Then

$$\sum_{i=1}^{n} \gamma_i(x_i, y_i) = -\log(P_{X^n|Y^n}(\mathbf{x}, \mathbf{y})) - H(X^n|Y^n) \ . \qquad\qquad (2)$$

5

Using Markov's inequality, for any $t > 0$,

$$\Pr_{\mathbf{x},\mathbf{y}}\Big[\sum_{i=1}^n \gamma_i(x_i, y_i) \geq n\delta\Big] = \Pr_{\mathbf{x},\mathbf{y}}\big[2^{t\sum_{i=1}^n \gamma_i(x_i,y_i)} \geq 2^{tn\delta}\big] \leq \frac{\mathrm{E}_{\mathbf{x},\mathbf{y}}\big[2^{t\sum_{i=1}^n \gamma_i(x_i,y_i)}\big]}{2^{tn\delta}} \;.$$

(3)

Moreover, because the pairs (x_i, y_i) are chosen independently,

$$\mathrm{E}_{\mathbf{x},\mathbf{y}}\Big[2^{t\sum_{i=1}^n \gamma_i(x_i,y_i)}\Big] = \mathrm{E}_{\mathbf{x},\mathbf{y}}\Big[\prod_{i=1}^n 2^{t\gamma_i(x_i,y_i)}\Big]$$

$$= \prod_{i=1}^n \mathrm{E}_{x_i,y_i}\big[2^{t\gamma_i(x_i,y_i)}\big] \leq \big(2^{\frac{1}{2}t^2 \log^2(|\mathcal{X}|+3)}\big)^n$$

where the inequality follows from Lemma 2, for any $|t| \leq \frac{1}{\log(|\mathcal{X}|+3)}$. Combining this with (3) gives

$$\Pr_{\mathbf{x},\mathbf{y}}\Big[\sum_{i=1}^n \gamma_i(x_i, y_i) \geq n\delta\Big] \leq 2^{\frac{1}{2}nt^2 \log^2(|\mathcal{X}|+3) - tn\delta} \;.$$

With $t := \frac{\delta}{\log^2(|\mathcal{X}|+3)}$ (note that $t \leq \frac{1}{\log(|\mathcal{X}|+3)}$ because $\delta \leq \log(|\mathcal{X}|)$), we conclude

$$\Pr_{\mathbf{x},\mathbf{y}}\Big[\sum_{i=1}^n \gamma_i(x_i, y_i) \geq n\delta\Big] \leq 2^{-\frac{n\delta^2}{2\log^2(|\mathcal{X}|+3)}} \;.$$

The first inequality of the theorem then follows from (2).

Similarly, if $t < 0$,

$$\Pr_{\mathbf{x},\mathbf{y}}\Big[\sum_{i=1}^n \gamma_i(x_i, y_i) \leq -n\delta\Big] = \Pr_{\mathbf{x},\mathbf{y}}\Big[2^{t\sum_{i=1}^n \gamma_i(x_i,y_i)} \geq 2^{-tn\delta}\Big] \leq \frac{\mathrm{E}_{\mathbf{x},\mathbf{y}}\big[2^{t\sum_{i=1}^n \gamma_i(x_i,y_i)}\big]}{2^{-tn\delta}} \;,$$

and thus

$$\Pr_{\mathbf{x},\mathbf{y}}\Big[\sum_{i=1}^n \gamma_i(x_i, y_i) \leq -n\delta\Big] \leq 2^{\frac{1}{2}nt^2 \log^2(|\mathcal{X}|+3) + tn\delta} \;.$$

The second inequality follows with $t := -\frac{\delta}{\log^2(|\mathcal{X}|+3)}$. $\qquad\qquad\square$

2.2 Asymptotic equality of smooth entropy and Shannon entropy

Proof (of Theorem 1). We first prove the bound on the max-entropy $H_{\max}^\varepsilon(X^n|Y^n)$. For any $\mathbf{y} \in \mathcal{Y}^n$ with $P_{Y^n}(\mathbf{y}) > 0$, let $\bar{\mathcal{X}}_{\mathbf{y}}$ be the set of all n-tuples $\mathbf{x} \in \mathcal{X}^n$ such that

$$-\log(P_{X^n|Y^n}(\mathbf{x}, \mathbf{y})) \leq H(X^n|Y^n) + n\delta \;.$$

Furthermore, let $P_{\bar{X}^n\bar{Y}^n}$ be the nonnegative function on $\mathcal{X}^n \times \mathcal{Y}^n$ defined by

$$P_{\bar{X}^n\bar{Y}^n}(\mathbf{x}, \mathbf{y}) = \begin{cases} P_{X^nY^n}(\mathbf{x}, \mathbf{y}) & \text{if } \mathbf{x} \in \bar{\mathcal{X}}_{\mathbf{y}} \\ 0 & \text{otherwise.} \end{cases}$$

(4)

We can assume without loss of generality that $\delta \leq \log(|\mathcal{X}|)$ (otherwise, the statement is trivial). Hence, by the first inequality of Theorem 2, $\Pr_{\mathbf{x},\mathbf{y}}[\mathbf{x} \notin \bar{\mathcal{X}}_{\mathbf{y}}] \leq \varepsilon$, which implies

$$\left\| P_{X^nY^n} - P_{\bar{X}^n\bar{Y}^n} \right\|_1 \leq \varepsilon . \tag{5}$$

For any fixed $\mathbf{y} := (y_1, \ldots, y_n) \in \mathcal{Y}^n$ with $P_{Y^n}(\mathbf{y}) > 0$,

$$1 \geq \sum_{\mathbf{x} \in \bar{\mathcal{X}}_{\mathbf{y}}} P_{X_1|Y_1}(x_1, y_1) \cdots P_{X_n|Y_n}(x_n, y_n) \geq |\bar{\mathcal{X}}_{\mathbf{y}}| 2^{-H(X^n|Y^n)-n\delta} ,$$

where the second inequality follows from the definition of the set $\bar{\mathcal{X}}_{\mathbf{y}}$. Consequently, we have $|\bar{\mathcal{X}}_{\mathbf{y}}| \leq 2^{H(X^n|Y^n)+n\delta}$. Moreover, by the definition of $P_{\bar{X}^n\bar{Y}^n}$, the support of the function $\mathbf{x} \mapsto P_{\bar{X}^n\bar{Y}^n}(\mathbf{x}, \mathbf{y})$ is contained in $\bar{\mathcal{X}}_{\mathbf{y}}$. By the definition of max-entropy and (5)

$$H_{\max}^\varepsilon(X^n|Y^n) \leq \max_{\mathbf{y} \in \mathcal{Y}^n} \log\left|\mathrm{supp}\, P_{\bar{X}^n\bar{Y}^n}(\cdot, \mathbf{y})\right| \leq H(X^n|Y^n) + n\delta .$$

To prove the bound on the min-entropy $H_{\min}^\varepsilon(P_{X^nY^n}|P_{Y^n})$, let $\bar{\mathcal{X}}_{\mathbf{y}}$, for any $\mathbf{y} \in \mathcal{Y}^n$ with $P_{Y^n}(\mathbf{y}) > 0$, be the set of n-tuples $\mathbf{x} \in \mathcal{X}^n$ such that

$$-\log(P_{X^n|Y^n}(\mathbf{x}, \mathbf{y})) \geq H(X^n|Y^n) - n\delta ,$$

and let again $P_{\bar{X}^n\bar{Y}^n}$ be defined by (4). By the second inequality of Theorem 2, $\Pr_{\mathbf{x},\mathbf{y}}[\mathbf{x} \notin \bar{\mathcal{X}}_{\mathbf{y}}] \leq \varepsilon$, which gives

$$\left\| P_{X^nY^n} - P_{\bar{X}^n\bar{Y}^n} \right\| \leq \varepsilon .$$

Thus, by the definition of min-entropy,

$$\begin{aligned}
H_{\min}^\varepsilon(X^n|Y^n) &\geq \min_{\mathbf{y} \in \mathrm{supp}\, P_{Y^n}} \min_{\mathbf{x} \in \mathcal{X}^n} \log \frac{P_{Y^n}(\mathbf{y})}{P_{\bar{X}^n\bar{Y}^n}(\mathbf{x}, \mathbf{y})} \\
&\geq H(X^n|Y^n) - n\delta.
\end{aligned}$$

where the second inequality follows from the definition of the set $\bar{\mathcal{X}}_{\mathbf{y}}$. $\qquad\square$

3   On the Tightness of the Bounds

In this section, we show that Theorems 1 and 2 are almost tight. For this, we construct one particular family of distributions for which these theorems cannot be strengthened much.

Consider the set $\mathcal{X} = \{0, \ldots, |\mathcal{X}| - 1\}$, where $|\mathcal{X}| \geq 3$. We set

$$P_X(x) := \begin{cases} \frac{1}{2} & x = 0 \\ \frac{1}{2(|\mathcal{X}|-1)} & \text{otherwise.} \end{cases} \tag{6}$$

An explicit calculation gives

$$H(X) = 1 + \frac{1}{2}\log(|\mathcal{X}| - 1). \tag{7}$$

For a tuple $\mathbf{x} := (x_1, \ldots, x_n)$ let $z(\mathbf{x}) := \big|\{i \mid x_i = 0\}\big|$ be the number of zeros in $\mathbf{x}$. Then,

$$(-\log(P_{X^n}(\mathbf{x}))) - nH(X) = z(\mathbf{x}) + (n - z(\mathbf{x}))(1 + \log(|\mathcal{X}| - 1)) - nH(X)$$
$$= \left(\frac{n}{2} - z(\mathbf{x})\right)\log(|\mathcal{X}| - 1). \tag{8}$$

Further, we note that $z(\mathbf{x})$ is binomially distributed with $p = \frac{1}{2}$.

We first show that Theorem 2 is almost tight. For this we use a lower bound on a partial sum over binomial coefficients which is given in Appendix B (Lemma 9).

Theorem 3. *For any $\mathcal{X}$ with $|\mathcal{X}| \geq 3$ there exists a distribution P_X over $\mathcal{X}$ such that for $n \geq 12$, $0 \leq \delta \leq \frac{\log(|\mathcal{X}|-1)}{12}$, and for $\mathbf{x}$ chosen according to $P_{X^n} := (P_X)^n$,*

$$\Pr_{\mathbf{x}}\big[-\log(P_{X^n}(\mathbf{x})) \geq n\big(H(X) + \delta\big)\big] > \frac{1}{110} 2^{-\frac{12n\delta^2}{\log^2(|\mathcal{X}|-1)}}$$

and, similarly,

$$\Pr_{\mathbf{x}}\big[-\log(P_{X^n}(\mathbf{x})) \leq n\big(H(X) - \delta\big)\big] > \frac{1}{110} 2^{-\frac{12n\delta^2}{\log^2(|\mathcal{X}|-1)}} .$$

Proof. Let P_X be the distribution defined by (6). We prove the second bound (the proof of the first bound is symmetric). From (8) we get

$$\Pr_{\mathbf{x}}\big[-\log(P_{X^n}(\mathbf{x})) \leq n\big(H(X) - \delta\big)\big] = \Pr_{\mathbf{x}}\left[\left(\frac{n}{2} - z(\mathbf{x})\right)\log(|\mathcal{X}| - 1) \leq -n\delta\right]$$
$$= \Pr_{\mathbf{x}}\left[\frac{n}{2} - z(\mathbf{x}) \leq -\frac{n\delta}{\log(|\mathcal{X}| - 1)}\right]$$
$$= \Pr_{\mathbf{x}}\left[z(\mathbf{x}) \geq \frac{n}{2} + \frac{n\delta}{\log(|\mathcal{X}| - 1)}\right].$$

Using Lemma 9 with $s := \lceil \frac{n\delta}{\log(|\mathcal{X}|-1)} \rceil$ (it is easy to check that the requirements

8

of Lemma 9 are satisfied) we get

$$\Pr_{\mathbf{x}}\left[z(\mathbf{x}) \geq \frac{n}{2} + \frac{n\delta}{\log(|\mathcal{X}|-1)}\right] \geq \sum_{k=\lceil\frac{n}{2}\rceil+s}^{\lceil\frac{n}{2}\rceil+2s-1} 2^{-n}\binom{n}{k}$$

$$> \frac{s}{2\sqrt{n}}e^{\frac{-8s^2}{n}}$$

$$\geq \frac{\sqrt{n}\delta}{2\log(|\mathcal{X}|-1)}e^{-\frac{8}{n}\left(\frac{n\delta}{\log(|\mathcal{X}|-1)}+1\right)^2}$$

$$= \frac{\sqrt{n}\delta}{2\log(|\mathcal{X}|-1)}e^{-\frac{8n\delta^2}{\log^2(|\mathcal{X}|-1)}-\frac{16\delta}{\log(|\mathcal{X}|-1)}-\frac{8}{n}}$$

$$\geq \frac{\sqrt{n}\delta}{2\log(|\mathcal{X}|-1)}e^{-\frac{8n\delta^2}{\log^2(|\mathcal{X}|-1)}-2}$$

$$\geq \frac{\sqrt{n}\delta}{16\log(|\mathcal{X}|-1)}2^{-\frac{12n\delta^2}{\log^2(|\mathcal{X}|-1)}}. \tag{9}$$

Fix now $\delta^* = \frac{\log(|\mathcal{X}|-1)}{4\sqrt{n}}$. We consider the cases $\delta \geq \delta^*$ and $\delta < \delta^*$ separately. First, in case $\delta \geq \delta^*$ equation (9) implies

$$\Pr_{\mathbf{x}}\left[z(\mathbf{x}) \geq \frac{n}{2} + \frac{n\delta}{\log(|\mathcal{X}|-1)}\right] > \frac{1}{64}2^{-\frac{12n\delta^2}{\log^2(|\mathcal{X}|-1)}}. \tag{10}$$

On the other hand, if $\delta < \delta^*$ we use (10) to get

$$\Pr_{\mathbf{x}}\left[z(\mathbf{x}) \geq \frac{n}{2} + \frac{n\delta}{\log(|\mathcal{X}|-1)}\right] \geq \Pr_{\mathbf{x}}\left[z(\mathbf{x}) \geq \frac{n}{2} + \frac{n\delta^*}{\log(|\mathcal{X}|-1)}\right] \geq \frac{1}{64}2^{-\frac{3}{4}} > \frac{1}{110}.$$

Since $1 \geq 2^{-\frac{12n\delta^2}{\log^2(|\mathcal{X}|-1)}}$ for all $\delta \geq 0$ this finishes the proof. $\qquad\square$

We now prove that Theorem 1 is almost tight.

Theorem 4. *For any $\mathcal{X}$ with $|\mathcal{X}| \geq 3$ there exists a distribution P_X over $\mathcal{X}$ such that for $n \geq 1200$ and $0 \leq \delta \leq \frac{\log(|\mathcal{X}|-1)}{480}$*

$$\frac{1}{n}H_{\max}^\varepsilon(X^n) \geq H(X) + \delta ,$$

$$\frac{1}{n}H_{\min}^\varepsilon(X^n) \leq H(X) - \delta ,$$

where $\varepsilon = \frac{1}{880}2^{-\frac{48n\delta^2}{\log^2(|\mathcal{X}|-1)}}$.

Proof. Again, let P_X be the distribution defined by (6). First, let $\mathcal{S} \subseteq \mathcal{X}^n$ be the set of values $\mathbf{x}$ whose probability is at most $P_{X^n}(\mathbf{x}) \leq 2^{-n(H(X)+2\delta+10/n)}$, and $\mathcal{T} \subseteq \mathcal{X}^n$ be the set of values whose probability is at least $P_{X^n}(\mathbf{x}) \geq$

$2^{-n(H(X)-2\delta-10/n)}$. For both $\mathcal{S}$ and $\mathcal{T}$, according to Theorem 3 (one can easily check that Theorem 3 can be applied for these parameters), the probability that $\mathbf{x}$ is in the set is at least

$$\frac{1}{110}2^{-\frac{12n(2\delta+10/n)^2}{\log^2(|\mathcal{X}|-1)}} = \frac{1}{110}2^{-\frac{12n(4\delta^2+40\delta/n+\frac{100}{n^2})}{\log^2(|\mathcal{X}|-1)}}$$

$$= \frac{1}{110}2^{-\frac{48n\delta^2}{\log^2(|\mathcal{X}|-1)}} \cdot 2^{-\frac{480\delta+\frac{1200}{n}}{\log^2(|\mathcal{X}|-1)}}$$

$$\geq \frac{1}{110}2^{-\frac{48n\delta^2}{\log^2(|\mathcal{X}|-1)}} \cdot 2^{-\frac{\log(|\mathcal{X}|-1)+1}{\log^2(|\mathcal{X}|-1)}}$$

$$\geq \frac{1}{440}2^{-\frac{48n\delta^2}{\log^2(|\mathcal{X}|-1)}} = 2\varepsilon. \tag{11}$$

We now prove the bound on $H_{\max}^\varepsilon$. If $Q \in \mathcal{B}^\varepsilon(P_{X^n})$ we have

$$\sum_{\mathbf{x}\in\mathcal{S}} Q(\mathbf{x}) \geq \sum_{\mathbf{x}\in\mathcal{S}} P_{X^n}(\mathbf{x}) - \sum_{\mathbf{x}\in\mathcal{S}} |P_{X^n}(\mathbf{x}) - Q(\mathbf{x})| \geq \varepsilon,$$

where the last inequality follows because (11) shows that $\sum_{\mathbf{x}\in\mathcal{S}} P_{X^n}(\mathbf{x})$ is at least 2ε, while $\sum_{\mathbf{x}\in\mathcal{S}} |P_{X^n}(\mathbf{x}) - Q(\mathbf{x})|$ is at most $\|P_{X^n} - Q\|_1 \leq \varepsilon$.

We can assume that for all $\mathbf{x} \in \mathcal{S}$ we have $Q(\mathbf{x}) \leq 2^{-n(H(X)+2\delta+10/n)}$ (otherwise we find a function $Q(\mathbf{x})$ which is closer to P_{X^n}, has the same support, and satisfies this) and thus this means that

$$|\operatorname{supp} Q| \geq \varepsilon\, 2^{n(H(X)+2\delta+10/n)},$$

i.e.,

$$\log(|\operatorname{supp} Q|) \geq n(H(X) + 2\delta + 10/n) - \log(1/\varepsilon).$$

Since

$$\log(1/\varepsilon) = \frac{48n\delta^2}{\log^2(|\mathcal{X}| - 1)} + \log(880) < n\delta + 10$$

we get that

$$H_{\max}^\varepsilon = \min_{Q_X \in \mathcal{B}^\varepsilon(P_{X^n})} \log|\operatorname{supp} Q_X|$$

$$\geq n(H(X) + 2\delta + 10/n) - n\delta - 10 = n(H(X) + \delta).$$

We now come to the bound on the min-entropy $H_{\min}^\varepsilon(X^n)$. If $Q \in \mathcal{B}^\varepsilon(P_{X^n})$ then $\varepsilon \geq \sum_{\mathbf{x}\in\mathcal{T}} |P_{X^n}(\mathbf{x}) - Q(\mathbf{x})| \geq \sum_{\mathbf{x}\in\mathcal{T}} P_{X^n}(\mathbf{x}) - Q(\mathbf{x})$ (where $\mathcal{T}$ is as defined above), and together with (11) this implies

$$\sum_{\mathbf{x}\in\mathcal{T}} Q(\mathbf{x}) \geq \sum_{\mathbf{x}\in\mathcal{T}} P_{X^n}(\mathbf{x}) - \varepsilon \geq \sum_{\mathbf{x}\in\mathcal{T}} \frac{1}{2}P_{X^n}(\mathbf{x}),$$

which implies that for some $\mathbf{x} \in \mathcal{T}$ we have $Q(\mathbf{x}) \geq 2^{-n(H(X)-\delta)}$. We therefore get

$$H_{\min}^\varepsilon(X^n) = \max_{Q\in\mathcal{B}^\varepsilon(P_{X^n})} \min_{\mathbf{x}\in\mathcal{X}^n} \log\left(\frac{1}{Q(\mathbf{x})}\right) \leq n(H(X) - \delta). \qquad \square$$

10

A Properties of the function r_t

This section lists some properties of the functions r_t defined by (1), i.e.,

$$r_t(z) = z^t - t \ln(z) - 1 .$$

These properties are used in Section 2.

Lemma 3. *For any $t \in \mathbb{R}$, the function r_t is monotonically increasing on the interval $[1, \infty)$.*

Proof. The first derivative of r_t is given by

$$\tfrac{d}{dz} r_t(z) = t z^{t-1} - \tfrac{t}{z} = \tfrac{t}{z}(z^t - 1) .$$

The assertion follows because the term on the right hand side is nonnegative for any $z \in [1, \infty)$. $\qquad\square$

Lemma 4. *For any $t \in \mathbb{R}$ and $z \in (0, \infty)$,*

$$r_t(z) \leq r_{|t|}(z + \tfrac{1}{z}) .$$

Proof. Observe first that $r_t(z) = r_{-t}(\tfrac{1}{z})$. It thus suffices to show that the statement holds for $t \geq 0$. If $z \geq 1$, the assertion follows directly from Lemma 3. For the case where $t \geq 0$ and $z < 1$, let $v := t \ln(1/z)$. Then $r_t(\tfrac{1}{z}) = e^v - v - 1$ and $r_t(z) = e^{-v} + v - 1$. Because $v \geq 0$, we have $e^v - e^{-v} \geq 2v$, which implies $r_t(z) \leq r_t(\tfrac{1}{z})$. The assertion then follows again from Lemma 3. $\qquad\square$

Lemma 5. *For any $t \in [-\tfrac{1}{2}, \tfrac{1}{2}]$, the function r_t is concave on the interval $[4, \infty]$.*

Proof. We show that $\tfrac{d^2}{dz^2} r_t(z) \leq 0$ for any $z \geq 4$. Because $\tfrac{d^2}{dz^2} r_t(z) = t(t-1)z^{t-2} + \tfrac{t}{z^2}$, this is equivalent to $t(1-t)z^t \geq t$. It thus suffices to verify that

$$z \geq \left(\frac{1}{1 - t} \right)^{\frac{1}{t}} ,$$

for any $z \geq 4$. If we substitute $s := \tfrac{1}{1-t}$ the right hand side of the above expression is $s^{\frac{s}{s-1}}$ (where $s \in [\tfrac{2}{3}, 2]$), whose derivate can be easily seen to be non-negative (we use $\ln(s) \leq s - 1$), and thus takes its maximum at $s = 2$, in which case it equals to 4. $\qquad\square$

Lemma 6. *For any $z \in [1, \infty)$ and $t \in [-\tfrac{1}{\log(z)}, \tfrac{1}{\log(z)}]$*

$$r_t(z) \leq \left(1 - \ln(2) \right) \log^2(z) t^2 .$$

11

Proof. Let $v := t \ln(z)$. Then

$$\frac{r_t(z)}{t^2} = \frac{e^{t\ln(z)} - t\ln(z) - 1}{t^2} = \frac{e^v - v - 1}{v^2} \ln^2(z) . \tag{12}$$

We first show that the term on the right hand side is monotonically increasing in v, that is,

$$\frac{d}{dv} \frac{e^v - v - 1}{v^2} = \frac{e^v - 1}{v^2} - 2\frac{e^v - v - 1}{v^3} = \frac{e^v + 1}{v^2} - \frac{2}{v}\frac{e^v - 1}{v^2} \geq 0 .$$

We multiply the last inequality with $v^2 e^{-v/2}$ on both sides and see that it is equivalent to

$$1 \geq \frac{2}{v}\frac{e^{v/2} - e^{-v/2}}{e^{v/2} + e^{-v/2}} ,$$

which holds because, for any $v \in \mathbb{R}$,

$$\left| \frac{e^{v/2} - e^{-v/2}}{e^{v/2} + e^{-v/2}} \right| = |\tanh(\tfrac{v}{2})| \leq \tfrac{|v|}{2} .$$

Hence, in order to find an upper bound on (12), it is sufficient to evaluate the right hand side of (12) for the maximum value of v. By assumption, we have $v \leq \ln(2)$, i.e.,

$$\frac{e^v - v - 1}{v^2} \ln^2(z) \leq \left(1 - \ln(2)\right) \log^2(z) ,$$

which concludes the proof. $\qquad\square$

B   Partial Sums over Binomial Coefficients

Let

$$B_p(k|n) := \binom{n}{k} p^k (1-p)^{n-k} \tag{13}$$

be the probability of obtaining k successes from n independent Bernoulli trials. We will also use the binary Kullback-Leibler distance $D(q\|p)$, which is defined for arbitrary $p, q \in [0, 1]$ by

$$D(q\|p) := q \log\left(\frac{q}{p}\right) + (1-q) \log\left(\frac{1-q}{1-p}\right). \tag{14}$$

Lemma 7. *For $p \geq \frac{1}{2}$, $\varepsilon \geq 0$, $p + \varepsilon < 1$*

$$D(p + \varepsilon \| p) \leq \frac{\varepsilon^2}{2\ln(2)p(1-p)}.$$

Proof. Define the function $f_p(\varepsilon) := D(p + \varepsilon \| p)$. Taylor's Theorem states that there exists a $\delta \in [0, \varepsilon]$ such that

$$D(p + \varepsilon \| p) = f_p(\varepsilon) = f_p(0) + f'_p(0)\varepsilon + f''_p(0)\frac{\varepsilon^2}{2} + f'''_p(\delta)\frac{\varepsilon^3}{6}. \tag{15}$$

Explicit calculation yields $f_p(0) = f'_p(0) = 0$ and $f''_p(0) = \frac{1}{\ln(2)p(1-p)}$. Also we get

$$f'''_p(\varepsilon) = \frac{2p + 2\varepsilon - 1}{(p + \varepsilon)^2(p + \varepsilon - 1)^2 \ln(2)},$$

which is positive for $\varepsilon > 0$. Together with (15) this gives the lemma. $\square$

Proposition 3 (Stirling's Approximation). *For any $n > 0$*

$$e^{\frac{1}{12n+1}} < \frac{n!\, e^n}{\sqrt{2\pi n}\, n^n} < e^{\frac{1}{12n}}. \tag{16}$$

In the following lemma we are only interested in the lower bound on $B_p(k|n)$. However, the upper bound comes for free.

Lemma 8. *For $0 < k < n$ and $p \in [0, 1]$*

$$e^{-\frac{1}{12k} - \frac{1}{12(n-k)}} < B_p(k|n)\sqrt{2\pi \frac{k(n-k)}{n}}\, 2^{nD(\frac{k}{n}\|p)} < 1.$$

Proof. We get

$$B_p(k|n)\sqrt{2\pi \frac{k(n-k)}{n}}\, 2^{nD(\frac{k}{n}\|p)}$$

$$= \underbrace{p^k(1-p)^{n-k}}_{2^{(k\log(p) + (n-k)\log(1-p))}} \frac{n!}{k!\,(n-k)!} \cdot \sqrt{2\pi \frac{k(n-k)}{n}} \cdot 2^{(k\log(\frac{k}{pn}) + (n-k)\log(\frac{n-k}{(1-p)n}))}$$

$$= \sqrt{2\pi \frac{k(n-k)}{n}} \cdot \frac{n!}{k!\,(n-k)!} \cdot 2^{k\log(\frac{k}{n}) + (n-k)\log(\frac{n-k}{n})}$$

$$= \sqrt{2\pi \frac{k(n-k)}{n}} \cdot \frac{n!}{k!\,(n-k)!} \cdot \left(\frac{k}{n}\right)^k \cdot \left(\frac{n-k}{n}\right)^{n-k}$$

$$= \frac{n!\, e^n}{\sqrt{2\pi n}\, n^n} \cdot \frac{\sqrt{2\pi k}\, k^k}{k!\, e^k} \cdot \frac{\sqrt{2\pi(n-k)}\,(n-k)^{n-k}}{(n-k)!\, e^{n-k}}. \tag{17}$$

Using (16) three times we obtain

$$B_p(k|n)\sqrt{2\pi \frac{k(n-k)}{n}}\, 2^{nD(\frac{k}{n}\|p)} > e^{\frac{1}{12n+1}} e^{-\frac{1}{12k}} e^{-\frac{1}{12(n-k)}} > e^{-\frac{1}{12k} - \frac{1}{12(n-k)}}.$$

Analogously (and since either $\frac{1}{12n} < \frac{1}{12k+1}$ or $\frac{1}{12n} < \frac{1}{12(n-k)+1}$)

$$B_p(k|n)\sqrt{2\pi \frac{k(n-k)}{n}}\, 2^{nD(\frac{k}{n}\|p)} < e^{\frac{1}{12n}} e^{-\frac{1}{12k+1}} e^{-\frac{1}{12(n-k)+1}} < 1. \qquad \square$$

Corollary 1. *For $p \in [\frac{1}{2}, 1]$, and $pn \leq k < n$:*

$$B_p(k|n) > e^{-\frac{1}{6(n-k)}} \sqrt{\frac{n}{2\pi k(n-k)}} \, e^{-n\frac{(\frac{k}{n}-p)^2}{2p(1-p)}}.$$

Proof. From Lemma 8 we get

$$B_p(k|n) > e^{-\frac{1}{12k} - \frac{1}{12(n-k)}} \sqrt{\frac{n}{2\pi k(n-k)}} \, 2^{-nD(\frac{k}{n}\|p)}$$

$$\geq e^{-\frac{1}{6(n-k)}} \sqrt{\frac{n}{2\pi k(n-k)}} \, 2^{-nD(\frac{k}{n}\|p)},$$

where we used $k \geq \frac{n}{2}$. Using the estimate in Lemma 7 concludes the proof. $\square$

Lemma 9. *Let $p \in [\frac{1}{2}, 1]$, $n, s \in \mathbb{N}$ such that $pn + 3s \leq n$. Then,*

$$\sum_{k=\lceil pn \rceil + s}^{\lceil pn \rceil + 2s - 1} B_p(k|n) > \frac{s}{2\sqrt{n}} e^{-\frac{2s^2}{np(1-p)}}.$$

Proof. Clearly, $\frac{n}{k(n-k)} \geq \frac{4}{n}$, for all values of n and k in the sum. Since $k < pn + 2s$ for all values in the above sum we get $\frac{k}{n} - p < \frac{2s}{n}$, and also we see that $n - k \geq s$. Using this together with Corollary 1 thus implies for all k of interest

$$B_p(k|n) > e^{-\frac{1}{6s}} \sqrt{\frac{2}{\pi n}} \, e^{-n\frac{(\frac{2s}{n})^2}{2p(1-p)}}$$

$$> \frac{1}{2\sqrt{n}} e^{-\frac{2s^2}{np(1-p)}}.$$

Since there are s summands we get the lemma. $\square$

References

[Che52]   H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, Dec. 1952.

[CT91]   T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. Wiley, New York, 1991.

[ILL89]   R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions (extended abstract). In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, pages 12–24, 1989.

[Ren05]   R. Renner. *Security of Quantum Key Distribution*. PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich, 2005. Also available at http://arxiv.org/abs/quant-ph/0512258.

[RW04]   R. Renner and S. Wolf. Smooth Rényi entropy and applications. In
          2004 IEEE International Symposium on Information Theory, page
          233, 2004.

[RW05]   R. Renner and S. Wolf. Simple and tight bounds for information
          reconciliation and privacy amplification. In *Advances in Cryptology
          — ASIACRYPT 2005*, volume 3788 of *Lecture Notes in Computer
          Science*, pages 199–216. Springer-Verlag, 2005.

[RWW06]   R. Renner, S. Wolf, and J. Wullschleger. The single-serving channel
          capacity. In *Proceedings of the 2006 IEEE International Symposium
          on Information Theory*, 2006.